



**Bussi, Javier; Marí, Gonzalo; Méndez Fernanda**

*Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística*

## **BOOTSTRAP RÁPIDO Y ROBUSTO PARA DATOS DEPENDIENTES: ESTIMADOR DE LA FUNCIÓN DE AUTOCORRELACIÓN EN MODELOS AR(1) CON MÚLTIPLES OUTLIERS**

### **Resumen**

La función de autocorrelación (FAC) es una herramienta fundamental en el análisis de series de tiempo lineales. La estimación muestral de la FAC es altamente sensible a la presencia de observaciones extremas. El objetivo del presente trabajo es comparar el estimador obtenido a través de una adaptación del método Bootstrap Rápido y Robusto (FRB según sus siglas en inglés) para estimar la FAC, con distintos estimadores propuestos en la literatura para datos dependientes. Esta comparación se realiza a través del sesgo medio y mediano y el Error Cuadrático Medio y Mediano. Cuatro de estos estimadores son variaciones basadas en la técnica Jackknife para series de tiempo con bloques móviles, tres estimadores contemplan por un lado al estimador clásico y por el otro a este mismo con observaciones truncadas al 2% y 5%. El restante es un estimador MG altamente robusto propuesto por Ma y Genton. Se comparan las estimaciones para el rezago de orden 1 de la FAC de un modelo AR(1) bajo distintos escenarios de simulación que proponían distintas longitudes de series ( $n=60,120,180$ ), distintos porcentajes (5%,10%) y ocurrencias en el tiempo (principio/medio/final) de la contaminación y distintos valores del parámetro del modelo ( $\phi=\pm 0,9; \pm 0,6, \pm 0,3$ ). El estimador basado en el método FRB resulta tener un comportamiento similar al estimador altamente robusto MG, y superior a todos los restantes estimadores, tanto desde el punto de vista del sesgo como del error cuadrático. Los estimadores basados en la técnica Jackknife resultaron tener comportamientos más pobres entre los estimadores robustos.

### **Abstract**

The autocorrelation function (ACF) is a fundamental tool in the analysis of linear time series. The sample estimate of the ACF is highly sensitive to the presense of outliers. The aim of this study is to compare an estimator of the ACF that results from an adaptation of the Fast and Robust Bootstrap Method (FRB), with different estimators for dependent data proposed in the literature. The mean and median bias and the mean and median Square Error (SE) of each estimator are used for the comparison. Four of the estimators compared are variations based on the Jackknife Method for time series with moving blocks. Three others consider the clasical estimator and two versions of it with truncated observations (2% and 5%). The remaining estimator is the highly robust estimator MG proposed by Ma and Genton. The estimates for the lag of order 1 of the ACF of an AR(1) model are compared under different simulation scenarios that involve different lengths for the series ( $n=60,120,180$ ), different percentages (5%,10%) and times of occurrence (beginning/middle/end) for the contamination and also different values for the parameter in the model ( $\phi=\pm 0,9; \pm 0,6, \pm 0,3$ ). The estimator based on the FRB and the MG estimator seem to have similar performances, which are superior to the performances of all the other estimators, not only from the bias but also from



the square error point of view. Estimators based on the Jackknife Method behave poorly among the robust estimators considered here.

**Palabras claves:** FUNCIÓN DE AUTOCORRELACIÓN, ESTIMADORES ROBUSTOS, JACKKNIFE, FAST ROBUST BOOTSTRAP

## INTRODUCCIÓN

La función de autocorrelación (FAC) es una herramienta fundamental en el análisis de series de tiempo lineales. Además de los fines descriptivos, es utilizada, entre otras muchas cuestiones, para la identificación del modelo. La estimación muestral de la FAC es altamente sensible a la presencia de observaciones extremas, generándose sesgos de distinto tipo. Varias alternativas robustas han sido propuestas para solucionar o atenuar el problema. El objetivo del presente trabajo es comparar el estimador obtenido a través de una adaptación del método Bootstrap Rápido y Robusto (FRB) para estimar la FAC, con distintos estimadores propuestos en la literatura para datos dependientes. Esta comparación se realiza a través del sesgo y el Error Cuadrático. Para ambas medidas utilizadas se computó el valor medio y mediano. Cuatro de los estimadores que toman parte en la comparación son variaciones basadas en la técnica Jackknife para series de tiempo con bloques móviles. Se presentan además, el estimador clásico de la FAC y dos variaciones del mismo que descartan en distinto porcentaje las observaciones mayores y menores de la serie, es decir, con 2% y 5% de observaciones truncadas. El restante es un estimador MG altamente robusto propuesto por Ma y Genton. Se compararon las estimaciones para el rezago de orden 1 de la FAC de un modelo AR(1) bajo distintos escenarios de simulación que proponían distintas longitudes de series ( $n=60, 120, 180$ ), distintos porcentajes (5% y 10%) y ocurrencias en el tiempo (principio/medio/final) de la contaminación y distintos valores del parámetro del modelo ( $\phi=\pm 0,9; \pm 0,6, \pm 0,3$ ).

## METODOLOGÍA

Sea un proceso estacionario  $z_t$ , se define la función de autocorrelación (FAC) de dos valores cualquiera de la variable  $z_t$  en los tiempos  $t$  y  $t + k$  como

$$\rho(t, t + k) = \rho(k) = \frac{\text{Cov}(z_t; z_{t+k})}{\sqrt{\text{Var}(z_t) \text{Var}(z_{t+k})}} = \frac{\gamma(k)}{\gamma(0)}, k = 0, \pm 1, \pm 2, \dots$$

A partir de una serie de tiempo observada  $z_1, z_2, \dots, z_n$ , se proponen diversos estimadores de la FAC. En primer lugar, se consideró la estimación muestral clásica por el método de los momentos (Est) la cual resulta sensible a la presencia de observaciones extremas. La misma se obtiene de la siguiente forma

$$\hat{\rho}_s(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}, \quad k = 0, 1, 2, \dots, (n-1)$$

donde  $\bar{z} = \frac{1}{n} \sum_{t=1}^n z_t$



Como ya se mencionó, el estimador Est resulta sensible a la aparición de valores considerados outliers debido a que se basa en estimadores no robustos. Por tal motivo, se presentan una serie de estimadores robustos presentes en la bibliografía. Se consideró un estimador robusto de la FAC que trunca un porcentaje de las observaciones más extremas. Se postularon dos valores para el truncamiento: 2% y 5% que dieron origen a los estimadores Trun2 y Trun5, respectivamente (Chan & Wei, 1992). Por otro lado, se presenta un estimador basado en un estimador de escala altamente robusto (MG) (Ma & Genton, 2000). Luego se consideraron estimadores basados en los métodos de replicaciones *Jackknife* y *Bootstrap* (Shao & Tu, 1995; Lahiri, 2003). Para el primero de ellos, cuatro estimadores basados en la técnica *Jackknife* para series de tiempo con bloques móviles en donde el primero de ellos utiliza la estimación muestral de la FAC (JackMB), los dos siguientes utilizan los estimadores robustos que truncan un porcentaje (2%, 5%) de las observaciones extremas (JackMB2, JackMB5 respectivamente) (Chambers, 2013), y el cuarto combina la estimación robusta MG y la técnica *Jackknife*, que inserta este estimador computándolo en cada réplica seleccionada (JackMB-MG). Por último se postuló el estimador a través del método *Bootstrap Rápido y Robusto* (FRB) (Salibian-Barrera, 2000; Salibian-Barrera & Zamar, 2002). Se partió de la estimación propuesta para los estimadores de regresión y se lo adaptó al caso de un modelo autorregresivo de primer orden AR(1). Para cada serie generada se obtuvo un estimador FRB replicando 1000 muestras bootstrap. Una vez estimado el parámetro correspondiente, el mismo fue utilizado en la estimación de la FAC.

A continuación se presenta brevemente, los estimadores propuestos.

#### *Estimador truncado de la función de autocorrelación*

Dada una serie de tiempo observada ordenada  $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$ , la función de autocorrelación muestral  $\alpha$ -recortada o truncada (Trun $\alpha$ ) está definida por:

$$\hat{\rho}_{T, \text{Trun}\alpha}(k) = \frac{\hat{\gamma}_{T, \alpha}(k)}{\hat{\gamma}_{T, \alpha}(0)}, \quad k = 0, 1, 2, \dots, (n-1)$$

donde

$$\hat{\gamma}_{T, \alpha}(k) = \frac{1}{\sum_{t=k+1}^n L_{t-k}^{(\alpha)} L_t^{(\alpha)}} \left\{ \sum_{t=k+1}^n (z_{t-k} - \bar{z}^{(\alpha)})(z_t - \bar{z}^{(\alpha)}) L_{t-k}^{(\alpha)} L_t^{(\alpha)} \right\}$$

con

$$\bar{z}^{(\alpha)} = \frac{\sum_{t=1}^n z_t L_t^{(\alpha)}}{\sum_{t=1}^n L_t^{(\alpha)}}$$

y

$$L_t^{(\alpha)} = \begin{cases} 0 & \text{si } z_t \leq z_{(g)} \text{ ó } z_t \geq z_{(n-g+1)}, \\ 1 & \text{en otro caso} \end{cases}$$



donde  $g$  es la parte entera de  $[\alpha n]$  y  $0 \leq \alpha < 0.5$ .

Para el presente trabajo se consideraron dos valores de  $\alpha$ , 2 y 5 por ciento.

#### *Estimador de escala altamente robusto de la función de autocorrelación*

Propuesto por Ma y Genton (2000), el estimador propuesto, a diferencia del estimador empírico, se basa en una medida de escala en lugar de una medida de posición. El estimador robusto viene dado por

$$\hat{\gamma}_{MG}(k) = 1/4 \left( Q_{n-k}^2(\mathbf{u} + \mathbf{v}) - Q_{n-k}^2(\mathbf{u} - \mathbf{v}) \right)$$

donde  $\mathbf{u} = (z_1, z_2, \dots, z_{n-k})$  y  $\mathbf{v} = (z_k, z_{k+1}, \dots, z_n)$  y  $Q_n(\mathbf{x})$  es un estimador de escala robusto propuesto por Rousseeuw y Croux (1992, 1993) que se calcula de la siguiente forma

$$Q_n(\mathbf{x}) = c \left\{ |z_i - z_j|; i < j \right\}_{(h)}$$

donde  $c$  es una constante elegida de forma tal que el estimador sea consistente ( $c=2.21191$  para la distribución normal),  $h = \left[ \left( \binom{n}{2} + 2 \right) / 4 \right] + 1$ ,  $\left\{ |z_i - z_j|; i < j \right\}_{(h)}$  representa la estadística de orden  $h$ -ésima de las diferencias absolutas entre pares de la serie observada.

#### *Estimador basados en el método Jackknife*

El método *Jackknife* fue desarrollado por Quenouille (1949) dentro del ámbito de las series de tiempo. El mismo consiste en la eliminación de una o varias observaciones de un conjunto de datos denominándose al conjunto resultante réplica, en la estimación del parámetro con la muestra reducida, y en repetir este procedimiento un número determinado de veces, obteniendo a partir del promedio de las estimaciones, un estimador del parámetro de interés.

Uno de los métodos de selección de réplicas es el denominado bloques móviles, donde, dado un bloque de longitud fija, el mismo se mueve a través de la serie de tiempo de a una observación por vez, calculando en cada movimiento el estimador correspondiente a la réplica luego de eliminar de la serie original el bloque móvil.

El método puede resumirse en los siguientes pasos

1. Definir un bloque de longitud  $n_M$ , lo que determina  $M$  bloques. Eliminar de la serie original las primeras  $n_M$  observaciones.
2. Con la muestra reducida estimar la FAC a partir de un método propuesto (*met*). Llamemos  $\hat{\rho}_{met(-1)}(k)$  a este estimador, donde  $(-1)$  indica que fue eliminado el primer bloque.
3. Mover el bloque una observación y repetir el paso 2.
4. Repetir el paso 3 hasta llegar al final de la serie



5. El estimador de la FAC a partir de este método es

$$\hat{\gamma}_{Jack}(k) = \frac{N}{M-1} \hat{\rho}_{met}(k) - \frac{(N-M+1)}{(M-1)} \frac{1}{M} \sum_{i=1}^M \hat{\rho}_{met(-i)}(k)$$

Como puede observarse, el método Jackknife necesita definir con qué método se realizarán las estimaciones en el punto 2. Para la presente investigación, se consideraron el método basado en los momentos (JackMB), los estimadores truncados con 2 y 5% de fracción de truncamiento (JackMB2 y JackMB5 respectivamente), y el método debido a Ma y Genton (JackMB-MG). En todos los casos se utilizó un tamaño de bloque de 12 observaciones bajo el supuesto de que se estaban considerando series mensuales. Si bien en este caso se supone que el proceso subyacente es un AR(1) estacionario, esto no siempre podría ser la situación. Se podría estar lidiando con series que presentan estacionalidad y por lo tanto es aconsejable reemplazar observaciones por aquellas que respeten la estructura debido a la misma. Otra alternativa posible sería considerar algún múltiplo del período estacional, en este caso  $i \cdot 12$  siendo  $i$  un número entero.

#### *Estimador basados en el método Bootstrap Rápido y Robusto (FRB)*

Un modelo AR(1) puede ser escrito de la siguiente forma  $z_t = \phi z_{t-1} + a_t$ , donde  $\phi$  es la función de autocorrelación para el rezago 1. Luego, la estimación de la misma, puede ser visto un problema de estimación del coeficiente de regresión en un modelo de regresión simple. Se considerará el método FRB para la estimación del mismo.

Se describe a continuación el método FRB en el caso de regresión lineal en general. Se considera el escenario donde se cuenta con variables explicativas aleatorias y donde se aplica el método a estimadores de regresión MM. Las observaciones  $\mathbf{x}_i$  no están prefijadas como en diseño de experimentos sino que son variables aleatorias que se observan conjuntamente con la variable dependiente  $y_i$ . Se describe brevemente a continuación el modelo lineal. Las observaciones  $\mathbf{x}_i$  tienen dimensión  $p \times 1$  donde se considerará el caso de  $p=2$ , es decir, una regresión lineal con una variable explicativa y una ordenada al origen. Se cuenta con  $n$  observaciones independientes e idénticamente distribuidas (i.i.d) de vectores aleatorios  $(y_i, \mathbf{z}_i)'$  con distribución común  $H$  donde  $\mathbf{x}_i = (1, \mathbf{z}_i)'$ .

Se considera el modelo lineal:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}_0 + \sigma_0 \varepsilon_i \quad i = 1, \dots, n$$

La situación ideal corresponde a suponer que  $y_i$  y  $\mathbf{z}_i$  son independientes, donde  $y_i \sim F_0$  siendo ésta alguna distribución simétrica (en general la normal estándar),  $\mathbf{z}_i \sim G_0$ ,  $(y_i, \mathbf{z}_i)' \sim H_0$ . Con el fin de considerar la presencia de observaciones atípicas (outliers) y otros desvíos con respecto al modelo clásico, se supone que la distribución  $H$  pertenece al entorno de contaminación:

$$H_\epsilon = \{H = (1 - \epsilon)H_0 + \epsilon H^*\}$$



donde  $0 \leq \epsilon \leq 1/2$  y  $H^*$  es una distribución cualquiera no especificada.

Los estimadores MM de regresión están basados en dos funciones de pérdida  $\rho_0$  y  $\rho_1$ , las cuales determinan la eficiencia y el punto de quiebre de la estimación, respectivamente.

El estimador MM,  $\hat{\beta}_n$ , satisface la ecuación:

$$\frac{1}{n} \sum_{i=1}^n \rho_1' \left( \frac{y_i - \mathbf{x}_i' \hat{\beta}_n}{\hat{\sigma}_n} \right) = 0$$

donde  $\hat{\sigma}_n$  es un estimador S de escala que minimiza el M estimador de escala  $\hat{\sigma}_n(\beta)$  definido en la ecuación:

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{y_i - \mathbf{x}_i' \beta}{\hat{\sigma}_n} \right) = 0.$$

La distribución asintótica de los estimadores MM ya ha sido estudiada el caso del modelo central paramétrico donde  $H = H_0$ . Sin embargo, esta situación no se presenta en situaciones donde se pretende utilizar estimadores MM altamente robustos. El método FRB descrito a continuación conduce a estimadores consistentes de la covariancia de  $\hat{\beta}_n$  bajo condiciones generales que incluye el caso en que  $H \in H_\epsilon$ .

Sea  $\tilde{\beta}_n$  el estimador de regresión S asociado a la ecuación

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{y_i - \mathbf{x}_i' \tilde{\beta}_n}{\hat{\sigma}_n} \right) = 0$$

Se desea realizar inferencias acerca del parámetro de regresión  $\beta_0$ . Utilizando el mismo criterio de "plug-in" del método BC propuesto por Efron (1979), se propone el siguiente método que produce un gran número de estimadores de regresión  $\hat{\beta}_n^*$  recalculados a partir de las muestras generadas. A partir de la función de distribución empírica de esta estadística se estima la distribución muestral del estimador de regresión  $\hat{\beta}_n$ . Para cada vector  $(y_i, \mathbf{z}_i)'$  perteneciente a la muestra se definen los residuos asociados con  $\hat{\beta}_n$  y  $\tilde{\beta}_n$ :

$$r_i = y_i - \mathbf{x}_i' \hat{\beta}_n$$

$$\tilde{r}_i = y_i - \mathbf{x}_i' \tilde{\beta}_n$$

Se debe notar que tanto  $\hat{\beta}_n$  como  $\hat{\sigma}_n$  pueden ser expresados como el resultado de un ajuste por mínimos cuadrados ponderados (MCP). Para  $i = 1, \dots, n$  se definen los pesos  $\omega_i$  y  $v_i$  como:

$$\omega_i = \rho_1'(r_i/\hat{\sigma}_n)/r_i$$



$$v_i = \frac{\hat{\sigma}_n \rho_0(\tilde{r}_i / \hat{\sigma}_n)}{nb \tilde{r}_i}$$

pudiendo entonces representarse los estimadores de la siguiente manera:

$$\hat{\beta}_n = \left[ \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i y_i$$

$$\hat{\sigma}_n = \sum_{i=1}^n v_i (y_i - \mathbf{x}_i' \hat{\beta}_n)$$

Si se considera la muestra bootstrap de las observaciones  $\{(y_i^*, \mathbf{x}_i^*)', i = 1, \dots, n\}$  y se definen las variables aleatorias:

$$\hat{\beta}_n^* = \left[ \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* \mathbf{x}_i^{*'} \right]^{-1} \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* y_i^*$$

$$\hat{\sigma}_n^* = \sum_{i=1}^n v_i^* (y_i^* - \mathbf{x}_i^{*'} \hat{\beta}_n)$$

donde:

$$\omega_i^* = \rho_1'(r_i^* / \hat{\sigma}_n) / r_i^*$$

$$v_i^* = \frac{\hat{\sigma}_n \rho_0(\tilde{r}_i^* / \hat{\sigma}_n)}{nb \tilde{r}_i^*}$$

$$r_i^* = y_i^* - \mathbf{x}_i^{*'} \hat{\beta}_n$$

$$\tilde{r}_i^* = y_i^* - \mathbf{x}_i^{*'} \tilde{\beta}_n,$$

para  $i = 1, \dots, n$ .

Es importante notar que los estimadores  $\hat{\beta}_n$ ,  $\tilde{\beta}_n$  y  $\hat{\sigma}_n$  no se recalculan para cada muestra bootstrap. Por lo tanto los estimadores recalculados pueden no reflejar la variabilidad del vector aleatorio  $(\tilde{\beta}_n, \hat{\sigma}_n)'$  debido a que los pesos  $\omega^*$  y  $v^*$  son calculados utilizando estimaciones fijas. Con el fin de ajustar los resultados a causa de esta situación, se aplica una corrección lineal a los estimadores  $\hat{\beta}_n^*$  y  $\hat{\sigma}_n^*$  recalculados y se los combina.

Sea entonces:



$$\mathbf{M}_n = \hat{\sigma}_n \left[ \sum_{i=1}^n \rho_1''(r_i/\hat{\sigma}_n, \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i'$$

$$\mathbf{d}_n = a_n^{-1} \left[ \sum_{i=1}^n \rho_1''(r_i/\hat{\sigma}_n, \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \rho_1''(r_i/\hat{\sigma}_n, \mathbf{x}_i) r_i \mathbf{x}_i$$

$$a_n = \frac{1}{nb} \left[ \sum_{i=1}^n \rho_0'(\tilde{r}_i/\hat{\sigma}_n) \tilde{r}_i/\hat{\sigma}_n \right]$$

El valor recalculado de  $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$  a través del bootstrap rápido está dado por:

$$\hat{\boldsymbol{\beta}}_n^{R*} - \hat{\boldsymbol{\beta}}_n = \mathbf{M}_n(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) + \mathbf{d}_n(\hat{\sigma}_n^* - \hat{\sigma}_n)$$

Es importante notar que al recalculer  $\hat{\boldsymbol{\beta}}_n^{R*} - \hat{\boldsymbol{\beta}}_n$  se resuelve el sistema de ecuaciones

$$\hat{\boldsymbol{\beta}}_n^* = \left[ \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* \mathbf{x}_i^{*'} \right]^{-1} \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* y_i^*$$

y se calcula el promedio ponderado

$$\hat{\sigma}_n = \sum_{i=1}^n v_i^* (y_i^* - \mathbf{x}_i^{*'} \hat{\boldsymbol{\beta}}_n)$$

Los factores de corrección  $\mathbf{M}_n$ ,  $\mathbf{d}_n$  y  $a_n$  surgen de resolver dos sistemas lineales y un promedio ponderado respectivamente, y se calculan solo una vez con la muestra completa. Para estimadores de regresión MM con  $\rho_0'$  redescendiente, los pesos  $\omega_i$  brindan al método estabilidad ante la presencia de outliers. Los puntos que estén alejados estarán asociados a pesos bajos en

$$\hat{\boldsymbol{\beta}}_n = \left[ \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i y_i$$

y en

$$\hat{\sigma}_n = \sum_{i=1}^n v_i (y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_n)$$





Los puntos atípicos extremos, es decir aquellos con residuos asociados muy grandes, recibirán pesos nulos y por lo tanto no tendrán efecto alguno sobre los coeficientes recalculados. Los pesos  $v_i$  utilizados en el recálculo de la escala son también decrecientes a medida que aumenta el valor absoluto de los residuos y en consecuencia las observaciones atípicas son menos influyentes también en el recálculo de  $\hat{\sigma}_n^*$ .

## RESULTADOS

Con el fin de comparar el estimador basado en el método Bootstrap Rápido y Robusto (FRB) con los demás estimadores presentados, se realizó un estudio de simulaciones donde se generaron 1000 series de longitud  $n$  igual a 60, 120 y 180, de procesos AR(1) ( $\phi = \pm 0,9; \pm 0,6; \pm 0,3$ ) contaminadas con múltiples outliers aditivos e innovativos (ver Anexo) en tres períodos distintos de las series generadas. Estos períodos fueron: al principio, en donde los outliers se asignaron aleatoriamente en las 20 primeras observaciones, al medio, asignándose las observaciones extremas en las 20 posiciones centrales y al final, en donde se asignaron aleatoriamente en las 20 observaciones finales. La cantidad de outliers distribuidos aleatoriamente correspondió a dos porcentajes distintos de contaminación: 5% y 10% del total de observaciones de la serie. Por ejemplo, para una serie de 180 observaciones se generaron 9 outliers en el caso de contaminación del 5% y estos fueron asignados aleatoriamente a 9 momentos de 20 disponibles, según lo explicado previamente, en cada serie en particular. Para detallar la comparación se cotejó la FAC teórica de los modelos simulados con cada una de las estimaciones propuestas. Se calcularon el sesgo medio y mediano y el error cuadrático (EC) medio y mediano de cada estimador sobre las 1000 replicaciones para el rezago  $k=1$ . El sesgo medio se calculó como la diferencia entre el promedio de los 1000 valores estimados y el valor teórico de la FAC según la siguiente fórmula:

$$\frac{\sum_{l=1}^{1000} \hat{\rho}_l}{1000} - \rho,$$

El sesgo mediano se computó como la mediana de las diferencias entre el valor estimado y el valor teórico.

$$\text{Mna}(\hat{\rho}_l - \rho) \quad l = 1, 2, \dots, 1000$$

El error cuadrático medio se calculó como el promedio de las diferencias al cuadrado entre el valor estimado y el valor teórico de la FAC.

$$\frac{\sum_{l=1}^{1000} (\hat{\rho}_l - \rho)^2}{1000}$$

El error cuadrático mediano se calculó como la mediana de las diferencias al cuadrado entre el valor estimado y el valor teórico de la FAC.

$$\text{Mna}(\hat{\rho}_l - \rho)^2 \quad l = 1, 2, \dots, 1000$$

Los outliers aditivos e innovativos se computaron en las cantidades y ubicaciones que se explicaron anteriormente, utilizando la magnitud  $\omega_A = \omega_I = 7 \times sd$  donde  $sd$  es el desvío estándar teórico del proceso AR (1) correspondiente.

A continuación, se detallan los resultados obtenidos a través de las simulaciones. En el caso de outliers aditivos, cuando el parámetro es positivo, todos los estimadores parecieran estar inclinados a subestimarlos mientras que ocurre lo contrario cuando el parámetro es negativo (Gráfico Nro. 1). Se puede observar que cualquiera sea el caso, el sesgo parece disminuir a

medida que aumenta el tamaño muestral, es decir, cuando se cuenta con series más largas (Gráfico Nro. 2). En general no existen diferencias entre el sesgo mediano y medio cualquiera sea el estimador considerado (Gráfico Nro. 3).

En el caso de outliers innovativos, a veces los estimadores subestiman o sobrestiman, independientemente del signo del parámetro e inclusive para un mismo escenario. Se puede observar también que los estimadores Jackknife, en todas sus versiones, tienen en general un rendimiento más pobre en el caso de outliers innovativos (Gráfico Nro. 4).

Gráfico N° 1 : Sesgo - 5% de Outliers Aditivos al Inicio

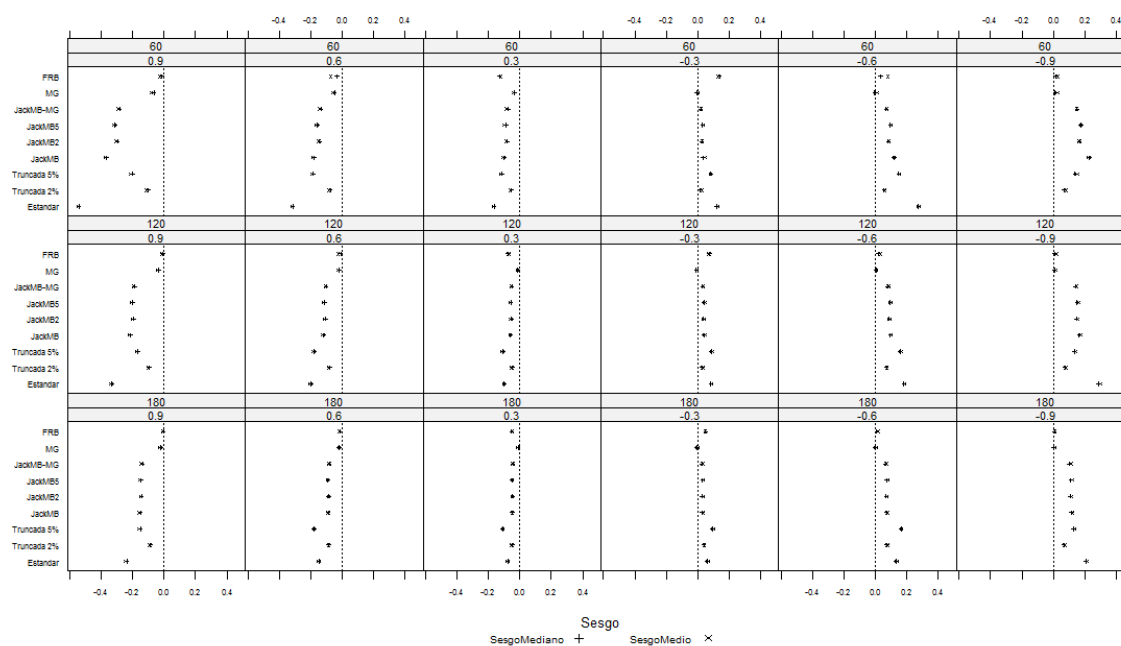


Gráfico N° 2 : Sesgo - 10% de Outliers Aditivos al Medio

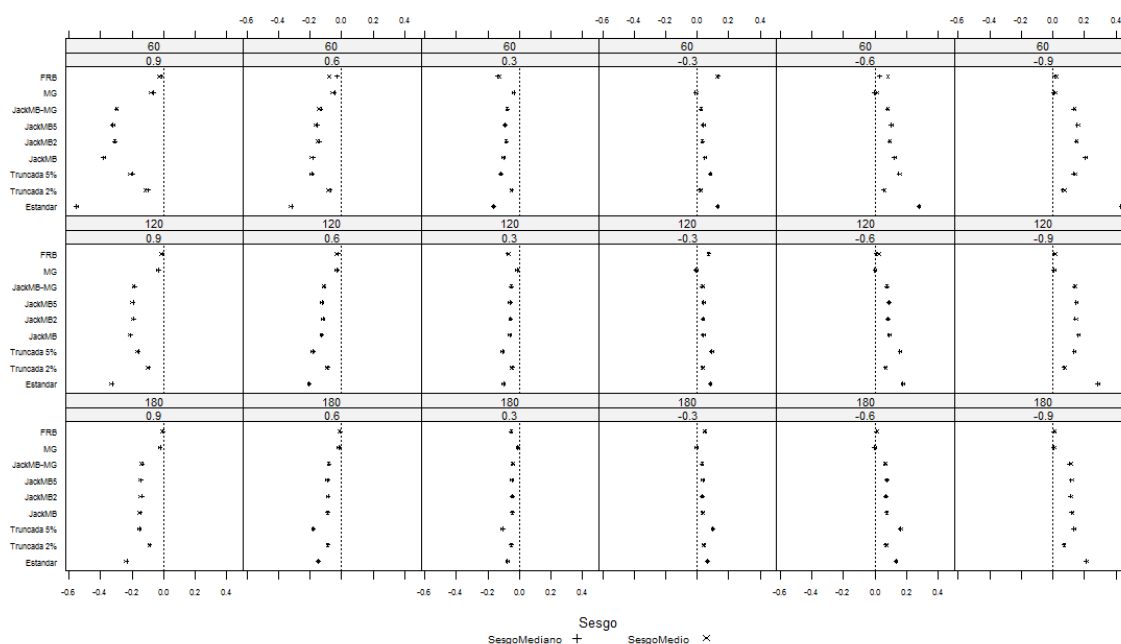


Gráfico N° 3 : Sesgo - 10% de Outliers Innovativos al Final

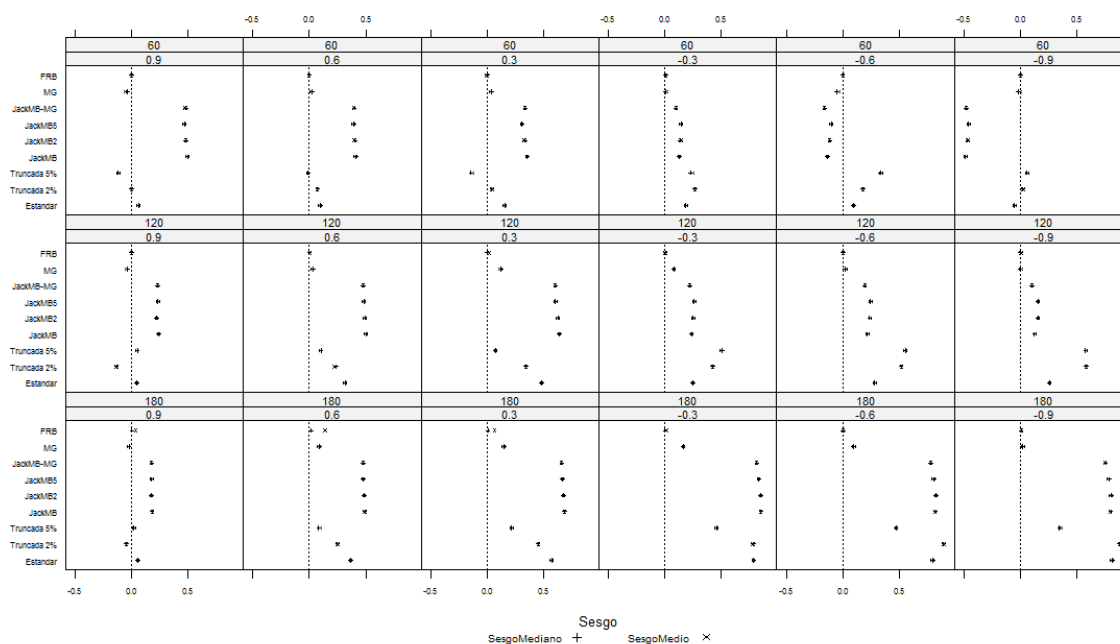
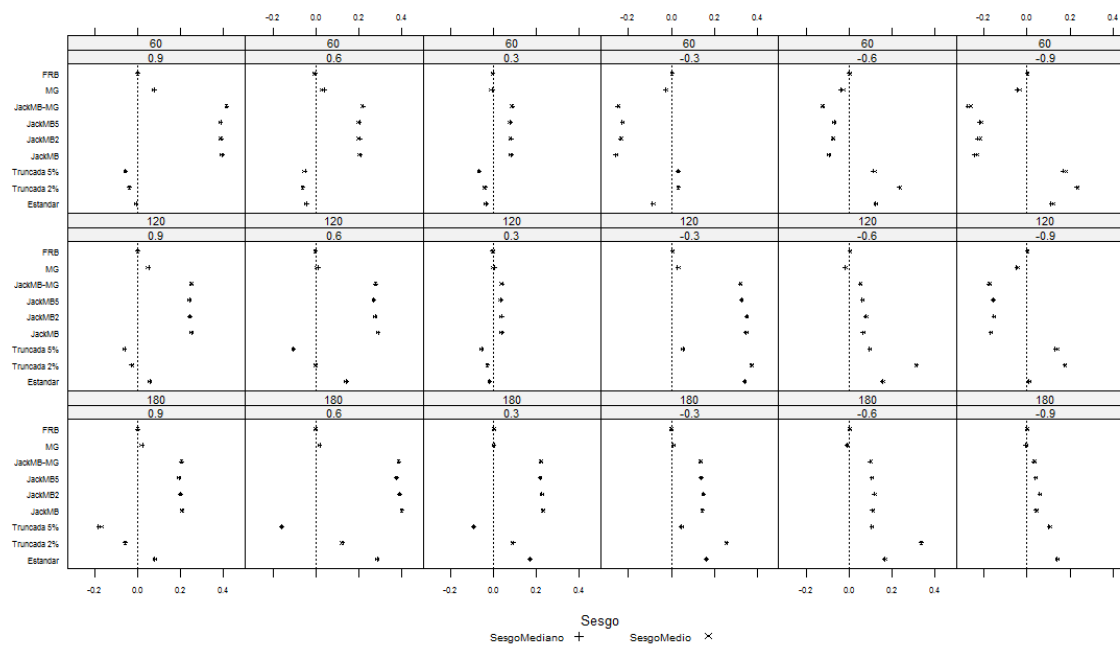


Gráfico N° 4 : Sesgo - 5% de Outliers Innovativos al Medio





En general, en cuanto al sesgo, el estimador FRB parece tener un buen desempeño, superado solo por el estimador MG, en algunas ocasiones, de manera leve, y principalmente en series cortas ( $n=60$ ) (Gráfico Nro. 2). Sin embargo, esto parecería compensarse ya que en algunas ocasiones el estimador FRB es superior al estimador MG, tanto en series cortas como largas (Gráfico Nro. 4).

En cuanto al EC medio y mediano, el estimador FRB tiene un buen desempeño con respecto al resto de los estimadores. Para el caso de outliers aditivos, el EC mediano es inferior al EC medio mostrando una asimetría en la distribución del mismo, pero esto ocurre solo para series cortas ( $n=60$ ) y casi exclusivamente para valores de  $\phi$  moderados (0,3 y 0,6). (Gráfico Nro. 5). En el caso de outliers innovativos, tanto el EC medio como el mediano coinciden en casi todos los casos, salvo en muy contadas ocasiones donde difieren levemente. (Gráfico Nro. 6).

Los estimadores Jackknife tienen el rendimiento más pobre entre los estimadores robustos en cuanto a EC medio y mediano, y también muestran asimetría en el caso de outliers aditivos, hecho que es más notorio en series cortas ( $n=60$ ) o intermedias ( $n=120$ ). (Gráfico Nro. 7).

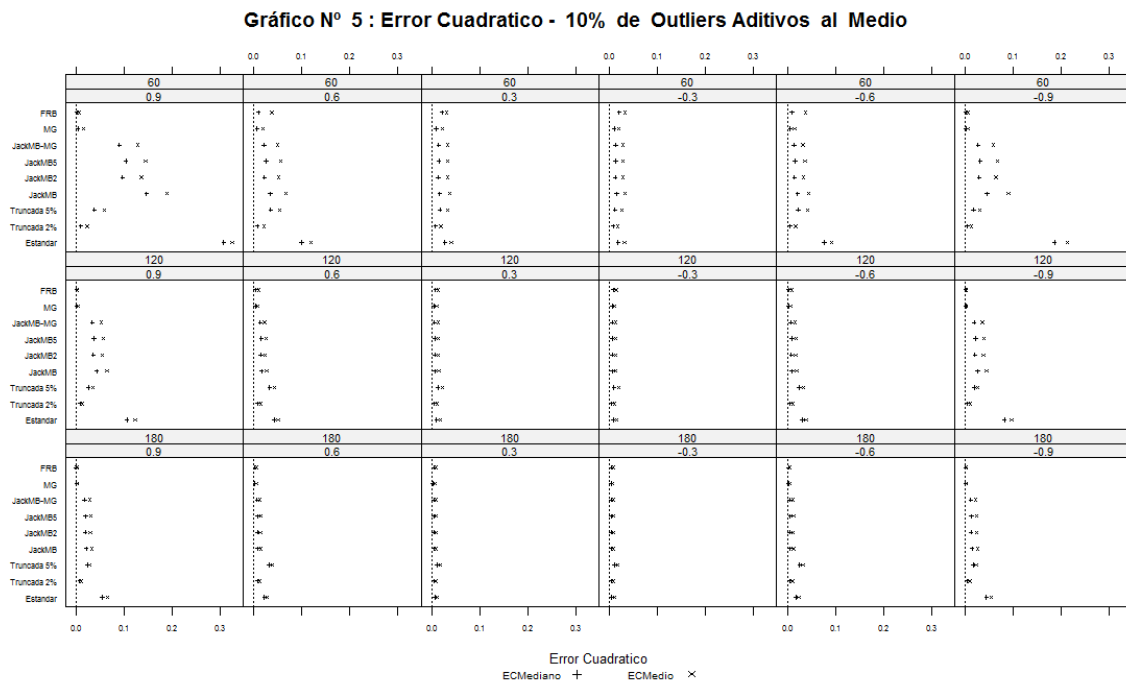




Gráfico N° 6 : Error Cuadrático - 10% de Outliers Innovativos al Medio

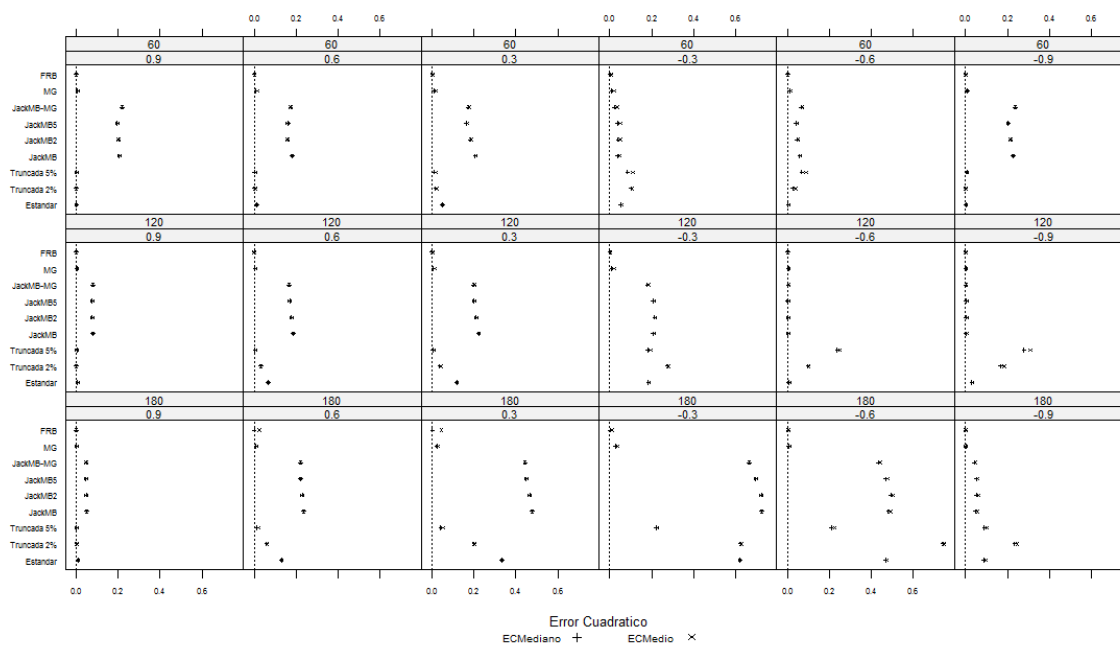
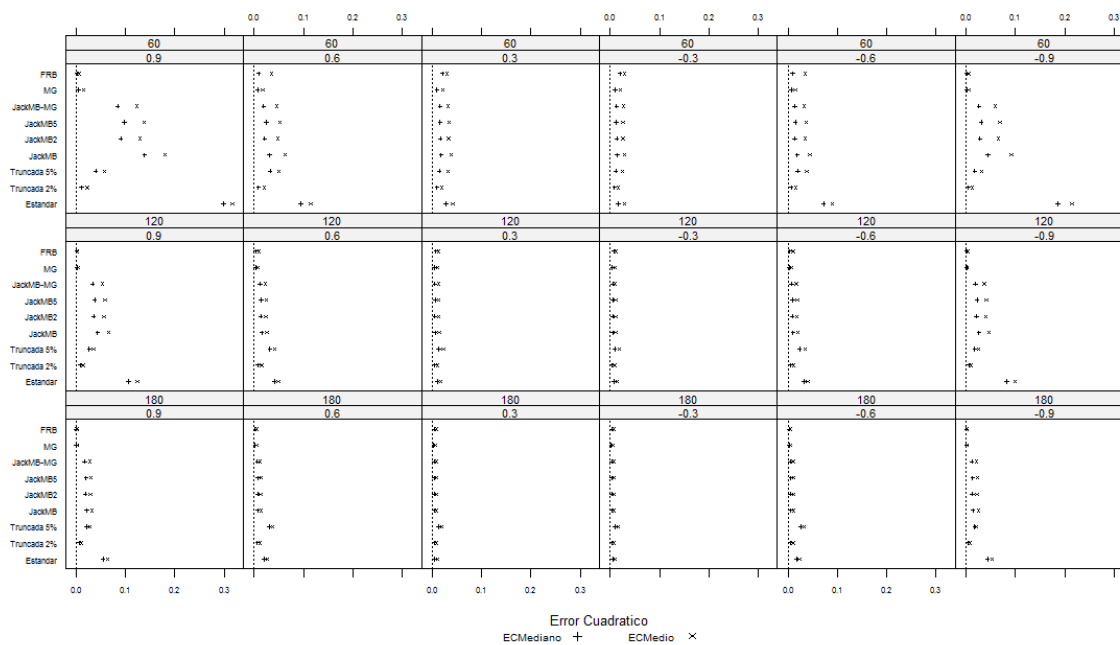


Gráfico N° 7 : Error Cuadrático - 5% de Outliers Aditivos al Final





## CONCLUSIONES

Los métodos de replicaciones para la estimación de la FAC de un modelo de serie de tiempo AR(1) presentaron resultados diversos cuando se presentan ciertos porcentajes de contaminación. El estimador basado en el método FRB parece ser una buena alternativa al estimador altamente robusto MG ya que resulta tener un comportamiento similar en los casos simulados. Los restantes estimadores parecieran ser inferiores tanto en sesgo como en error cuadrático medio y mediano, en particular los estimadores Jackknife.

## DISCUSIÓN

En trabajos anteriores se había vislumbrado el potencial del estimador FRB para datos dependientes en procesos AR(1) ante la presencia de alguna observación aberrante, ya sea aditiva o innovativa (Bussi, Marí & Méndez, 2015). Continuando con las propuestas postuladas ante los resultados prometedores obtenidos, se extendió el análisis a una mayor cantidad de contaminación y a escenarios donde la misma estuviese en distintas secciones de la serie, con el fin de someter el estimador a un desafío mayor. El estimador FRB produjo buenos resultados para datos dependientes, y en su naturaleza está presente el remuestrear a partir de los residuos obtenidos al computar un estimador robusto del parámetro correspondiente a la FAC. Este proceder resulta estar en concordancia con lo sugerido en la bibliografía dedicada a los métodos Bootstrap para datos dependientes, lo que refuerza los resultados obtenidos aquí donde el estimador FRB podría considerarse un estimador alternativo tan bueno como el de MG.

Si bien las simulaciones están acotadas a ciertos valores del proceso AR involucrado y a ciertas longitudes de series de tiempo, los resultados se presentan como muy alentadores, ya que las funciones de autocorrelación de procesos ARMA más generales podrían calcularse a partir de la estimación de funciones de autocorrelación de procesos AR infinitos.

Queda pendiente para futuros trabajos, explorar y comparar el rendimiento de los métodos de replicaciones para estimar la FAC en el caso de modelos MA puros y procesos ARMA, manteniendo los escenarios de contaminación presentados en este trabajo.



## REFERENCIAS BIBLIOGRÁFICAS

- Bonifazi, F.** (2015). *Estimación robusta de la función de autocorrelación*. Tesina de grado de la Licenciatura en Estadística, Facultad de Ciencias Económicas y Estadística, UNR.
- Bonifazi, F.; Méndez, F.** (2014). Estimación robusta de la función de autocorrelación. *Decimoctavas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística*.
- Box, G. E. P.; Jenkins, D. A.** (1976). *Time Series Analysis and Control*, 2<sup>nd</sup> edition. Holden-Day.
- Box, G. E. P.; Pierce, D. A.** (1970). Distribution of residual autocorrelations in autorregresive-integrated moving average time series models. *J. American Statistical Association*, 65, 1509-1526.
- Buishand, T.A.; Beersma, J.J.** (1993). Jackknife Tests for Differences in Autocorrelation between Climate Time Series. *Journal of Climate*, 1, 6, 2490-2495.
- Bussi, J.; Marí, G.; Méndez, F.** (2015). Estimación de la Función de Autocorrelación en modelos AR(1) con métodos de replicaciones. *Vigésimas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística*.
- Chambers, M. J.** (2013). Jackknife Estimation of Stationary Autoregressive Models. *Journal of Econometrics*, 171, Issue 1, January, 142–157
- Chan, W.; Wei, W.** (1992). A comparison of some estimators of time series autocorrelations. *Computational Statistics & Data Analysis*, 14, 146-163.
- Dürre, A.; Fried, R.; Liboschik, T.** (2014). *Robust estimation of (partial) autocorrelation*. Discussion Paper.
- Efron, B., Tibshirani, R.** (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Efron, B.** (1979). Bootstrap Methods: another look at the Jackknife. *Annals of Statistics*, 7, 1-26.
- Lahiri, S. N.** (2003). *Resampling Methods for Dependent Data*. Springer-Verlag
- Ma, Y.; Genton, M.** (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis*, 21, N°6, 663-684.
- Maronna, R.A.; Martin, R.D.; Yohai, V.J.** (2006). *Robust Statistics: Theory and Methods*. John Wiley and Sons.
- Quenouille, M.H.** (1949). Approximate tests of correlation in time series. Mathematical Proceedings of the Cambridge Philosophical Society, 45, 483-484



- Rousseeuw, P. J.; Croux, C.** (1992). Explicit scale estimators with high breakdown point. *L1- Statistical Analysis and Related Methods*, 77-92.
- Rousseeuw, P. J.; Croux, C.** (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273-83.
- Salibian-Barrera, M.** (2000). *Contributions to the theory of robust inference*. Ph.D. thesis, Dept. Statist., Univ. British Columbia, Vancouver.
- Salibian-Barrera, M., Van Aelst S., Willems G.** (2006). PCA Based on Multivariate MM-Estimators with Fast and Robust Bootstrap. *Journal of the American Statistical Association*, 101, 1198-1211.
- Salibian-Barrera, M., Zamar, R. H.** (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, 30, 556-582.
- Shao, J., Tu, D.** (1995). *The Jackknife and Bootstrap*. Springer-Verlag.
- Van Aelst, S., Willems, G.** (2005). Multivariate regression S-estimators for robust estimation and inference. *Statistica Sinica*, 15, 981-1001.
- Van Aelst, S., Willems, G.** (2013). Fast and robust bootstrap for multivariate inference: The R package FRB. *Journal of Statistical Software*, 53 (3), 1–32. URL: <http://www.jstatsoft.org/v53/i03/>.





## ANEXO

### -Outlier aditivo

Un *outlier* aditivo (OA) corresponde a un error externo o un cambio exógeno de los valores observados de la serie de tiempo en un instante determinado. Entonces se dice que ha ocurrido un OA en el instante  $T$ ,  $1 \leq T \leq n$ , sobre una serie temporal si el valor de la serie se genera en ese instante de manera distinta al resto. El modelo que sigue la serie observada  $z_t$  afectada por un OA en el momento  $t = T$  es:

$$\begin{aligned} z_t &= \begin{cases} y_t, & t \neq T \\ y_t + \omega_A, & t = T \end{cases} \\ &= y_t + \omega_A I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)} a_t + \omega_A I_t^{(T)} \end{aligned}$$

donde  $y_t$  es la serie no contaminada por atípicos que se supone sigue un modelo ARMA,  $\omega_A$  es la magnitud del efecto del atípico sobre la serie e  $I_t^{(T)}$  es una variable indicadora que representa la presencia o ausencia de *outliers* en el momento  $T$ , es decir  $I_t^{(T)} = 0, t \neq T$ ;  $I_t^{(T)} = 1, t = T$ .

### -Outlier innovativo

Los *outliers* innovativos (OI) son específicos de situaciones dinámicas, pueden ser generados por errores internos o cambios endógenos.

El modelo para la serie que sufre un OI de magnitud  $\omega_I$  en el instante  $t = T$  es:

$$\begin{aligned} z_t &= \begin{cases} y_t, & t < T \\ y_t + \frac{\theta(B)}{\phi(B)} \omega_I, & t = T + j, \quad j \geq 0 \end{cases} \\ &= y_t + \frac{\theta(B)}{\phi(B)} \omega_I I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)} (a_t + \omega_I I_t^{(T)}) \end{aligned}$$

donde  $y_t$  es la serie no contaminada por atípicos que se supone sigue un modelo ARMA,  $\omega_I$  es la magnitud del efecto del atípico sobre la serie e  $I_t^{(T)}$  es una variable indicadora que representa la presencia o ausencia de *outliers* en el momento  $T$ , es decir  $I_t^{(T)} = 0, t \neq T$ ;  $I_t^{(T)} = 1, t = T$ .